SafeSwim:
Overhead vision to help people see better

Steve Mann, Christina Mann, Faraz Sadrzadeh-Afsharazar, Zhao Lu, Jiten Saini, Jacqueline Lian
MannLab Canada, Toronto, Ontario, Canada, M5T 1G5

***Abstract*—Traditionally video surveillance consists of overhead cameras to help authorities such as lifeguards or security professionals monitor individuals. This monitoring, if in service of the monitored individuals, serves them only indirectly. More recently, a new kind of sensing called "sousveillance" has emerged, as a form of sensing by, and in direct service to, individuals. In this paper, we explore an overhead camera (e.g. aerial drone-based videography) to help create a form of real-time seeing aid, with specific application to help one or more individual swimmers see better and navigate, e.g. to avoid collisions with other swimmers (or the walls) in pools, as well as navigating in lakes, rivers, and other bodies of open water. Additionally the system can be used to teach swimming (e.g. learn by seeing from a birds-eye-view) as well as to gamify swim-based activities. SafeSwim (and sousveillance in general) is a new form of "seeing" in which those being watched have direct visual access to the live video stream. It is a direct departure from traditional surveillance where authorities such as police or lifeguards watch individuals without giving them direct access to a live video stream.**

***Index Terms*—Wearable computing, Wearables, Sensing, Mobile computing, Swimming, Drones, Sousveillance, Machine vision.**

I. Introduction

Wearable computing [1] combined with wearable AI [2] provides a new way of sensing and undertanding the enviroment around us. Wearable sensing technologies can create new and extended human sensing capabilities which work similarly to sensory organs. This concept is explored in Fig 1. The human mind may be regarded as one or more feedback loops. The human may similarly be regarded as a feedback loop between the mind and body, i.e. efferent nerves carry signals from the mind to the body, and afferent nerves carry signals from the body to the mind [3], thus completing a closed loop feedback control system [4], [5], shown in leftmost Fig. 1.

Humanistic Intelligence (HI) [6], also known as Wearable AI [2], is a framework for wearable computers and other forms of closed-loop feedback between a human and machine(s).

Technologies like shoes, wearable computers, bicycles, and automobiles are technologies that "become part of us." This symbiosis between human and machine is often referred to as a "humachine" [7], "cyborg" [5], bionic, or "augmented human". Humanistic technologies typically consist of sensors in the machine, and our human senses, forming a feedback loop, as shown in the center part of Fig. 1. The example shown here is a Vuzix "Smart Swim" eyeglass or other similar augmented swimglass. After wearing an eyeglass for a long period of

Authors thank Vuzix (Paul Travers) for equipment donation, and members of SwimOP™ (the online "Swim at Ontario Place" group we created), including Zhao Lu and Trevor McKee, for help with beach cleanup efforts.

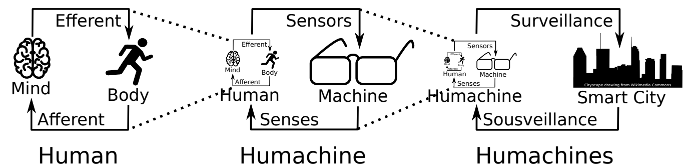

Fig. 1: Self, technology, and society as closed-loop feedback systems (Humanistic Intelligence). Note the fractal (self-similar) nature of humanistic intelligence, society, and the smart city, etc..

time, we adapt to seeing the world in this new way, and the technology becomes, in some sense, "part of us", through the long-term HI feedback loop [2].

In a given space or context, we often have more than one humachine, e.g. we might have 2 or more swimmers wearing the "Smart Swim" or other augmented reality eyeglass. Thus we need to think of Minsky et al.'s concept of the "Society of Intelligent Veillance" in which HI (wearable AI) applies across a community [8].

As we design and build smart pools, "Smart Beaches", smart buildings, smart streets, and smart cities, we need to also consider the idea of "smart people", i.e. putting "intelligence" (sensors and computation) on people, not just buildings and lampposts and the skies above them. In this way, humachines are more than just a nexus of human and machine, they form a new society and a new concept of technology-and-society. Therefore, HI feedback loops can be understood and classified according to the following nested taxonomy:

- **Self and technology** (e.g. combined human+machine);
- **Self and society** (interaction between humans, humachines, etc.), which includes self and technology;
- **Self and the environment** (interaction between the augmented human and the natural or built environment, e.g. cyborg-city interaction), which includes self, technology, and society.

This human-centered taxonomy considers the sensory information available at three key levels as shown in Fig 1). The third layer, "Self and the environment", becomes especially clear when we finally grant individuals a birds-eye-view of large bodies of open water, where an individual swimmer can see not only their stroke (i.e. use our system as a way to learn how to swim), but also the water conditions overall, such as algae blooms, floating garbage in some places, etc.. This insight gives an individual swimmer a new way to navigate.

For example, he or she can choose to swim along the path-of-least-pollution, or along the path-of-maximum-safety.

Pollution or safety or other disiderata/undisiderata can be overlaid in an augmented reality as a field, e.g. a danger-field or safety-field, such as to guide a swimmer using this new birds-eye-view, thus contextualizing sousveillance at its highest level – "Self and the environment".

This calls for a new way of thinking, in which smart pools, smart beaches, and smart cities provide direct, and immediate feedback to individual citizens, hence promoting the living quality of the end-users in the system.

A. *Safety and security tree*

To understand our new way of thinking and contrast it with traditional surveillance, we construct a mental model based on a tree. We can think of surveillance (security, etc.) as forming the trunk of a tree with various branches such as (1) crime prevention, (2) operational efficiency (e.g. so lifeguards can watch swimmers and see how big crowds are getting, and thus know where to deploy additional staff), and (3) accident prevention (e.g. automatic computer-assisted drowning detection), etc..

This model matches what one typically finds in smart city planning.

We propose, also, that there be roots for the tree, and that the roots form various aspects of sousveillance (inverse surveillance), such as to support personal safety, individual health and well-being, etc.

The "Securitree" is composed of two main sections, with the parts of the tree above the ground representing surveillance and the parts of tree below the ground representing sousveillance. We regard security as the trunk of a tree having surveillance branches like public safety, operational efficiency, and crime prevention. There are surveillance sub-branches, e.g. crime prevention has sub-branches such as robbery prevention, violence prevention, ... The tree is rooted in sousveillance, with roots like personal safety, "suicurity" (corruption prevention), personal efficiency, and individual's health. Sousveillance has sub-branches, e.g. individual health has sub-branches such as physical, affective, mental, (which itself has sub-roots VMP, seeing aid, . . .), as shown in Fig 2.

B. *EyeTap*

The EyeTap is a wearable device worn in front of one or both eyes, acting as if the eye itself is a camera. A display overlaying computer generated imagery can be viewed from the right eye. The left eye is unobstructed. Similar to the EyeTap, our SafeSwim™ system allows the user to be fully immersed in their surroundings. The user is able to see unobstructed through one eye, while their other eye has a display providing real-time safety-oriented situational awareness as demonstrated in Fig 7 from video processed from an overhead DJI Mavic Mini drone.

C. *Immersive/Submersive ("Mersive") VR*

Fully immersive/submersive (underwater) VR has been previously explored through "immersive multimedia" (underwater interactive virtual reality) [9] and VR float tanks [10], [11] using the Mersivity™ underwater VR headset (Fig 3).

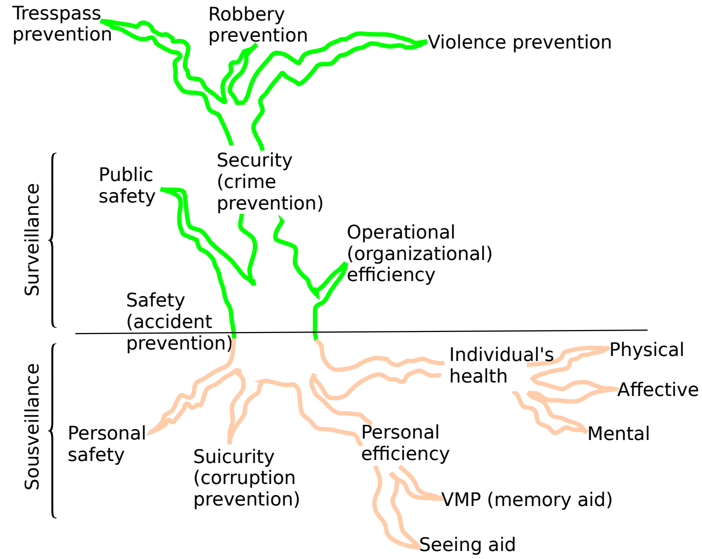

Fig. 2: The "Securitree"(Security Tree) model

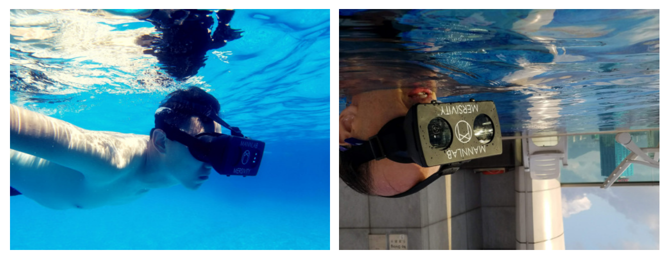

Fig. 3: Underwater virtual reality has been a well-established practice over the past 20 years or so. It provides a full immersive/submersive experience in which it is very easy to suspend disbelief while floating in a virtual world.

D. *Background*

The tracking of swimmers is a well-researched field. One of the state-of-the art swimmer tracking algorithms uses non-linear joint transform and colour histogram analysis to accurately detect the locations of the swimmers. Such a hybrid approach is needed for the reliable detection of the swimmers. In such an environment, the rapid movements of swimmers, bodily rotations, as well as water splashes can hinder the reproducibility of the detection [12], [13].

II. Methodology

A. *Proposed system architecture*

In a swimmer's context, in order to close the HI sensory feedback loop, some form of sensory technology is required. We are proposing to use a "sousveillance" (inverse surveillance) drone to capture a bird's eye view of the swimmer and make it available to the swimmer in real-time (this immediate availability to the end-user is what makes it sousveillance rather than surveillance). The swimmer is equipped with a

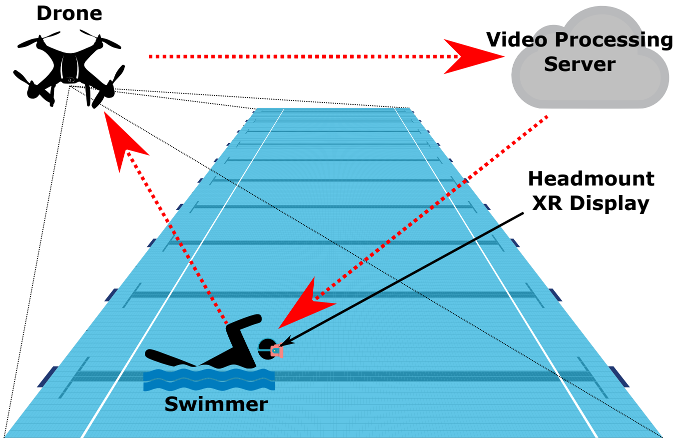

Fig. 4: The SafeSwim apparatus summary: an overhead drone oversees the whereabouts, swimming path, and the surroundings of the swimmer and computes and displays useful information to the swimmer, as a means of aiding the swimmer to see better.

head-mount mediated reality device capable of displaying the drone footage to the swimmer. Other useful information such as the swimmer's whereabouts, surroundings, and trajectory are then proposed to be overlaid onto the mentioned footage. A pictorial representation of our proposal is displayed in Fig. 4.

B. State of the art

The downside of the mentioned swimmer detection method is its lack of computational simplicity that would allow it to run in real-time. In order to make the detection algorithm run in real-time, solutions such as classical thresholding or Convolutional Neural Network methods are more appropriate. In order to run a classical thresholding algorithm, a distinctly coloured object was made present in frame. In order to realize this goal, the swimmer that was intended to be highlighted in the drone footage (the swimmer of interest) wore a brightly and distinctly colored (neon orange) swimsuit to separate the swimmer from the surrounding background in the frame. See Fig.5 displays the swimmer of interest wearing neon orange swimming trunks having unique color that was not present anywhere else in the swimming area.

Convolutional Neural Nets (CNN) were also implemented in parallel with the classical thresholding method for tracking the swimmer's path. This is further discussed in the computational methodology section.

C. Physical Experiment

We performed multiple experiments at multiple beaches and public pools in the city. In each experiment, the "User" would wear a distinct brightly colored swimsuit such as a pair of neon orange swimming shorts to facilitate the upcoming measures of post-processing the video footage.

While swimming, the "User" would try to use the image overlaid from the head-up display to guide his directions in

Fig. 5: Through use of neon orange swimming trunks, the targeted swimmer can be more quickly detected through classical thresholding with less delay (lag) than more advanced tensor-flow methods. Note the orange pylons around the pool also trigger the thresholding, so it is required to ignore this color when it occurs outside the bounding box of the pool.

swimming. Meanwhile, a drone would be capturing a video stream from the swimmer. The captured footage was then streamed onto the swimmer's MR wearable. Fig 6 better demonstrates the experimental setup. The way this was done was by pointing a smartphone camera at the drone's remote controller, which displayed the drone's live footage. Therefore, the mentioned was done as a compromising alternative. A light shroud shown was constructed to be able to effectively capture the screen of the drone controller in bright sunny days. In the future we will construct our own special-purpose sousveillance drone and write our own software to control the drone so that it can provide realtime chromakey video feedback with overlays with no noticable delay.

D. Video Processing

The videos were simultaneously recorded on the SD card that is available on the DJI drone, as well as streamed real-time to the controller cellphone. The videos were divided into 30 second segments to run post analysis. Each frame was then independently processed using equation 1. In this equation, "I" represents the input video, "C" is the threshold, and "Output" is the binary mask produced from the operation. The "C" variable had to be manually adjusted for each footage, due to the different lighting conditions present in each drone video. This variable was set in such a way that resulted in

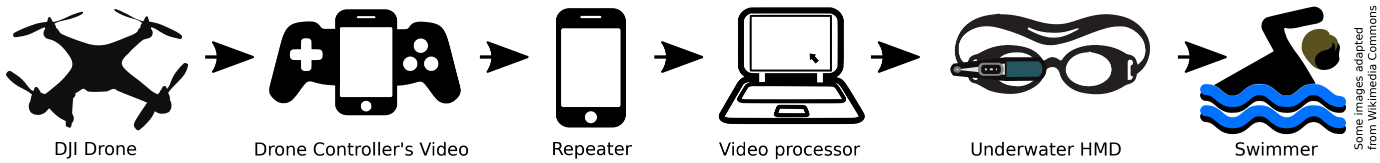

Fig. 6: Realtime signal flow path: iPhone on drone controller with EasyCast to Samsung S8 as repeater, to laptop computer for processing, to Vuzix SmartSwim (underwater head-mounted display) worn by swimmer.

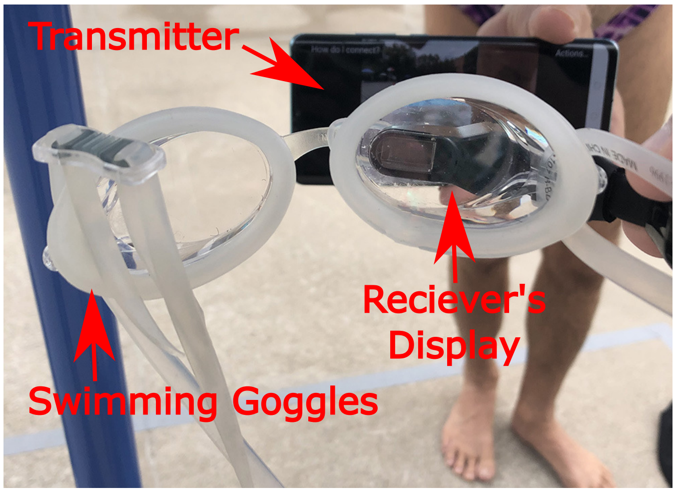

Fig. 7: Custom android apps for sending video feed from transmitter to receiver.

the minimum number of false positives, as well as the correct identification of the swimmer's neon swimming trunks.

$$Output = (I(x, y, t)_R - I(x, y, t)_G - I(x, y, t)_B) > C \quad (1)$$

For giving an experience of MR, we live-streamed the camera feed from the drone to the right eye of the swimmer wearing Vuzix Smart Swim. To achieve this, we used the real-time streaming network protocol (RTSP). We used two custom android apps - one acted as a transmitter and the other one as a receiver. The transmitter acted as an internet protocol(IP) camera and we live-streamed over a local network created via a mobile hotspot. The mobile hotspot also increased the wireless transmission range of the whole system by positioning it about midway between the transmitting camera and the swimmer (e.g. right at the shore of the beach, or right at the fence around the pool).

The receiver app was installed on the smart swim glasses which was running on an Android operating system. The receiver app was fit to the screen size of the Smart Swim glass. We used multithreading to handle incoming video data on the Smart Swim glass to reduce the delay and blockage of the main thread of the app in order to optimize the responsiveness of the overall experience. With the vibrant high-definition (HD) waterproof screen, the swimmer was able to see the live stream with clarity while swimming providing constant feedback of the environment ahead.

From the recorded video data, random frames of the pool were extracted. Each frame was labeled with the swimmers as persons for further training on a YOLOv3 network. It was required since normal YOLOv3 was not able to detect persons in the pool from the bird's eye view. We used transfer learning on the YOLOv3 network to detect humans in the pool from the top perspective. For this, we used labelImg to label the persons and then we trained our network over Google Colab GPU.

III. Tests and Results

After running the classical thresholding algorithm, red objects in the frame were able to be identified and highlighted. Fig 8,9,10 displays screenshots of the drone footage before and after being fed into the classical thresholding algorithm, described in equation 1 with other orange objects in the frame.

With the video data collected over several experiments, random frames were used to train a custom object detection model for detecting humans in the bird's eye view as shown in the figure, Fig 12. The best mean average precision (mAP) that we achieved was 0.7144.

Running a YOLOv3 model on HD frames will reduce the frame-rate processed, leading to delay in the feedback of the swimmer and reducing the size of the images, diminishing the size of the humans in it, and making it difficult for the model to detect. Smaller object detection is one of the current research areas in computer vision. One of the solutions is to use the focal loss as the loss function while training the model. Also, once the model is able to detect all the humans in the pool, the swimmer can choose to get feedback on his/her path and can also get notification based on the distance to other swimmers (e.g. early warning of possible collision) as shown in the figure, Fig 11.

IV. Other related applications

The proposed personal (sousveillance-based) computer vision system has many uses beyond merely navigation. It also opens up possibilities for "citizen science". The term "Open Science" was coined by Mann in 1998 cite Open Science, Open Source and R by Andy Wills, Linux Journal, February 19, 2019, and is a sousveillant-systems approach to science as may be done by ordinary individuals in their day-to-day life. Various sensors make it possible for large databases to be built up to study emergent phenomena.

Another use-case is in activities like citizen-driven beach cleanup where ordinary people make the world a better place. Here for example we observed and mapped the location of

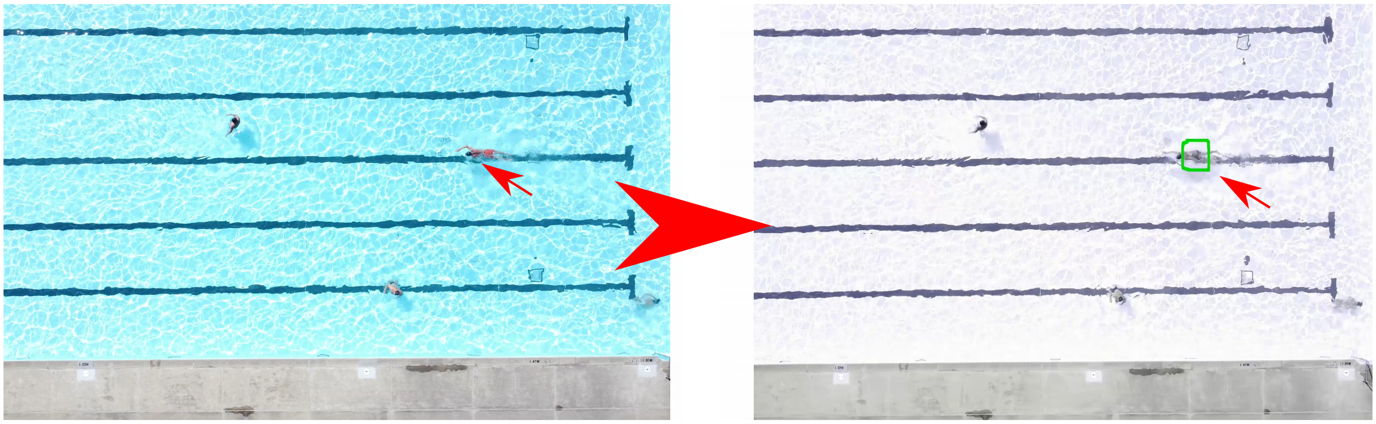

Fig. 8: Screenshots of drone footage before and after being fed into the classical thresholding swimmer detection algorithm. The red arrow highlights the swimmer of interest.

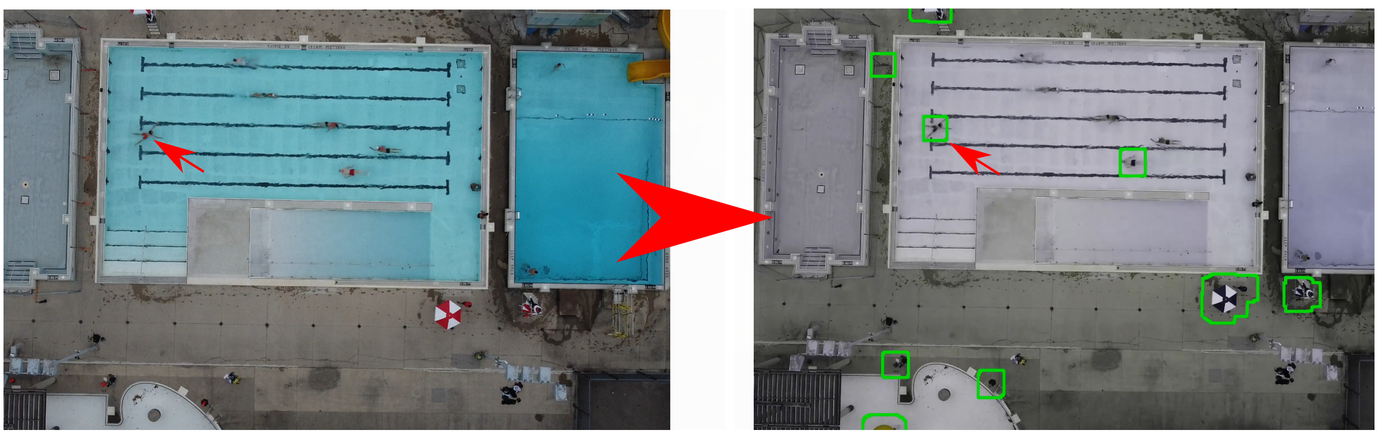

Fig. 9: Screenshots of drone footage before and after being fed into the classical thresholding swimmer detection algorithm. The red arrow highlights the swimmer of interest.

Fig. 10: Screenshots of drone footage before and after being fed into the classical thresholding swimmer detection algorithm. The red arrow highlights the swimmer of interest.

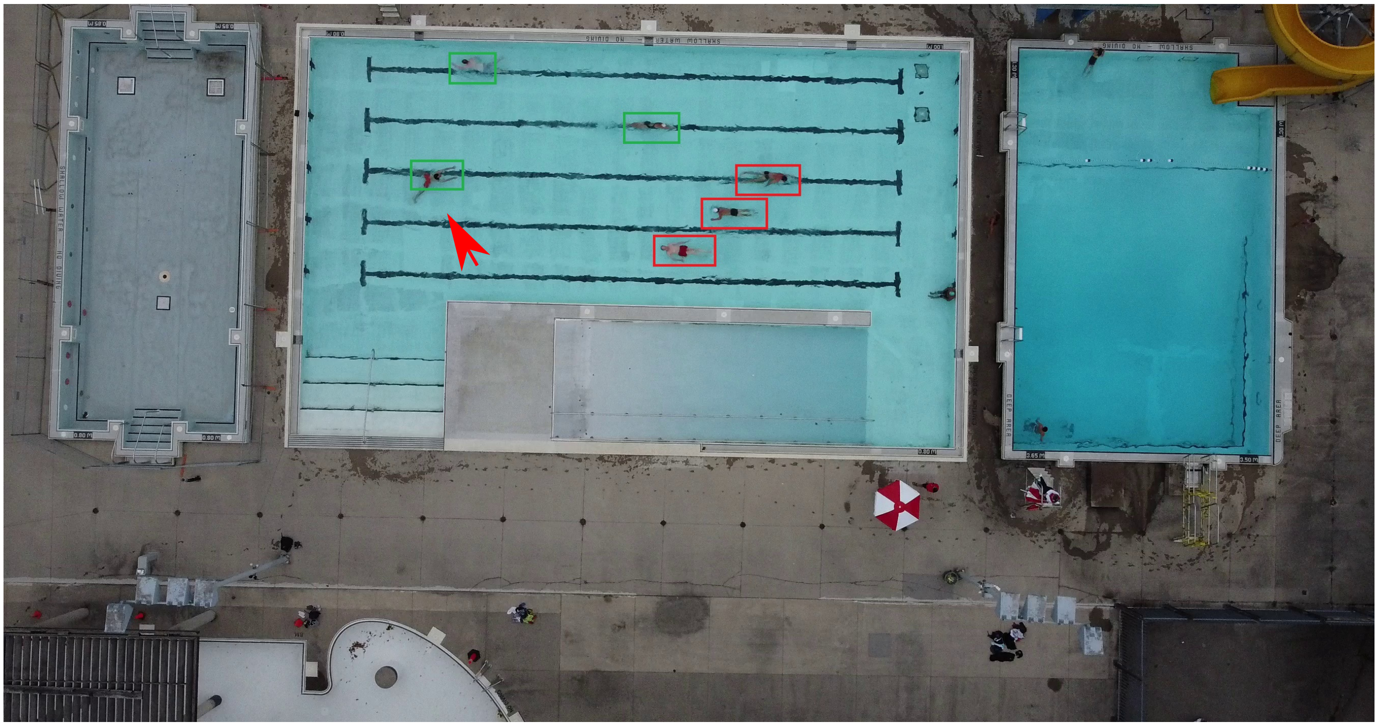

Fig. 11: The swimmer can get real time update of his/her distance with other swimmers. The closer ones shown in red. Blue arrow points to the swimmer of interests

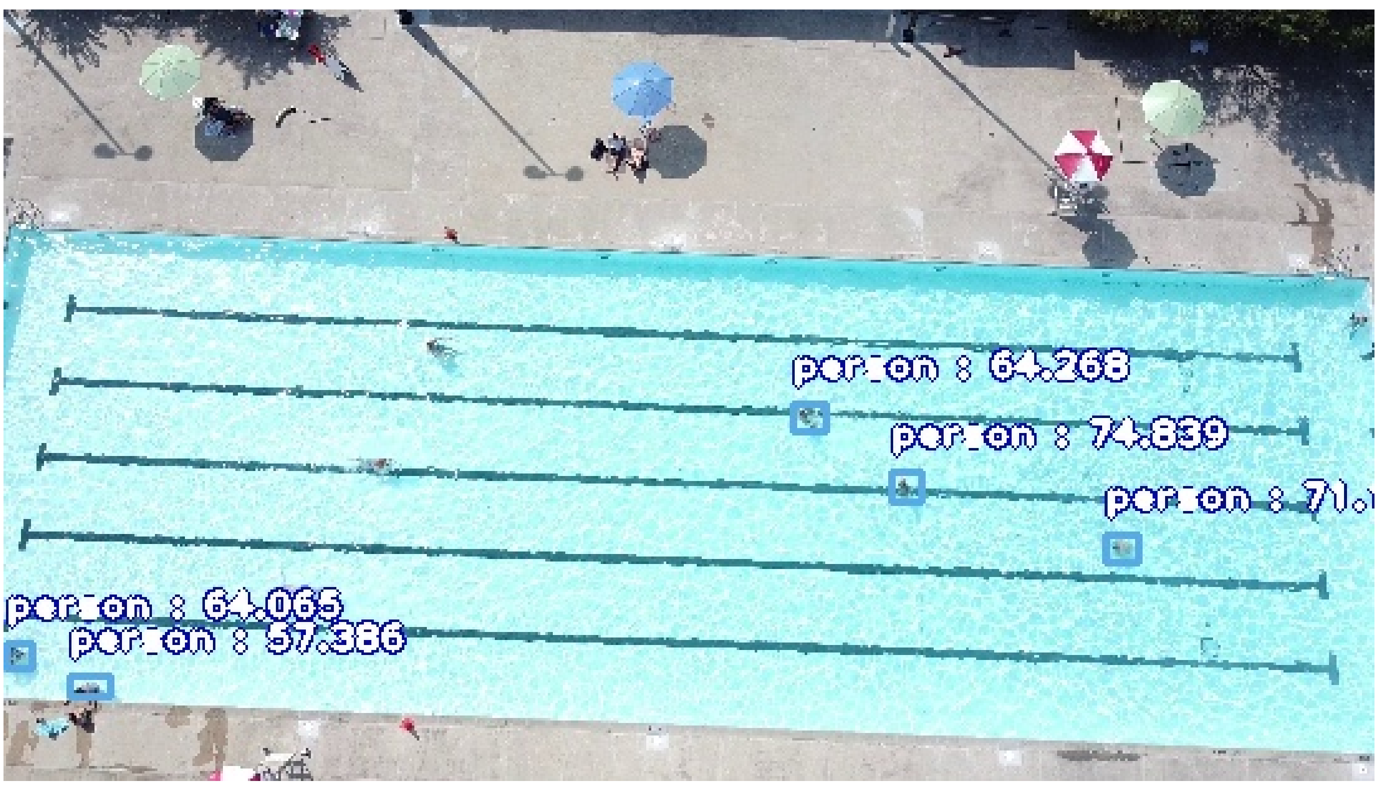

Fig. 12: Detection of humans from the bird's eye view.

large quantities of sharp metal underwater objects along the beach that represented a safety hazard. Our beach cleanup effort consisted of locating hazardous objects and removing them from the beach. In one example a large rusty metal railing with a sharp jagged metal base was tipped over on its side such that the sharp metal base was facing upwards on the bottom of the lake in a shallow area where swimmers would bump into it. One of the swimmers (S. Mann) was able to note its location from the overhead drone feed, and waited until a calm day (2020 November 9, wind speed 5km/h) to swim out and tie a long rope to it. A large number of volunteers on land pulled on the rope while swimmers S. Mann and T. McKee eased it off the bottom of the lake and guided it out. Many other sharp jagged metal hazards were removed in a similar way using the augmented reality capability of our system to assist in beach cleanup. See Fig. 13 and Fig. 14.

Fig. 13: Large quantities of sharp jagged pieces of rusty metal were removed once they were located and identified with SafeSwim. Author S. Mann swam out with a rope and, diving down, tied the rope to a large jagged metal railing.

V. Conclusion

We proposed a taxonomy and wearable computer system based on this taxonomy to help people sense and understand the world around them on the levels of (A) Self-and-Technology, (B) Self-and-Society, and (C) Self-and-Environment. The proposed framework and the functional system helps us understand the often blurry lines between the three levels of the taxonomy while meaningfully enhancing the wearer's senses.

Fig. 14: Once the rope was tied, a number of volunteers assisted in pulling the railing out.

Fig. 15: Virtual Reality waterball (Mersivity headset).

VI. Future Work

We have an ongoing project entitled "Equiveillance, Covidized Surveillance..." funded by the McLuhan Program in Culture and Technology to explore the socio-political elements of Weartech™. As a precursor to this project, consider the familiar waterball we often see in pools and lakes (Fig 15), as a form of social-distancer on land (Fig 16). We (S. Mann, C. Pierce, C. Tong, and C. Mann) have been exploring the social-commentary of social-distancing. See Fig. 17. Along these lines, SafeSwim may be regarded as part of the transition from surveillance and environmentalism to veillance and vironmentalism on land or in water.

References

[1] S. Mann, "Wearable computing: A first step toward personal imaging," *IEEE Computer*, vol. 30, no. 2, pp. 25–32, 1997.

[2] S. Mann, L.-T. Cheng, J. Robinson, K. Sumi, T. Nishida, S. Matsushita, Ömer Faruk Özer, O. Özün, C. Öncel Tüzel, V. Atalay, A. E. Çetin, J. Anhalt, A. Smailagic, D. P. Siewiorek, F. Gemperle, D. Salber, S. Weber, J. Beck, J. Jennings, and D. A. Ross, "Wearable ai," *Intelligent Systems, IEEE*, vol. 16, no. 3, pp. 1–53, May/June 2001.

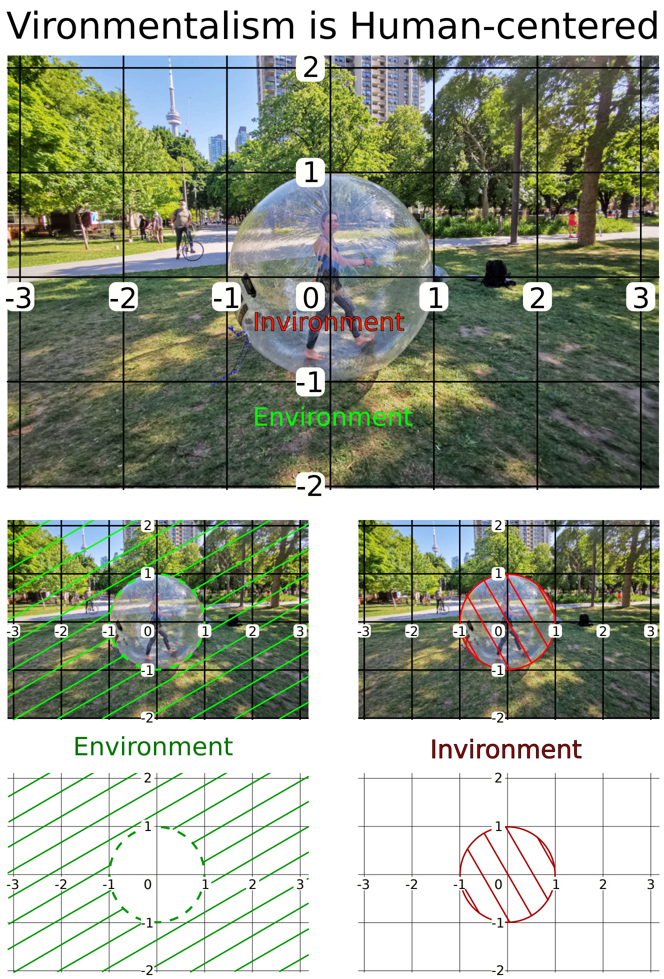

Fig. 16: "Vironment/Social-Bubble", situates the familiar "waterball" as social commentary on social-distancing. "Environment" is that which sourrounds us. "Inviroment" is us and our "social bubble". "Vironmentalism" emphasizes the boundary between us and our surroundings.

Fig. 17: Vironment as performance art and permanent outdoor sculpture.

[3] K. Müller, B. Ebner, and V. Hömberg, "Maturation of fastest afferent and efferent central and peripheral pathways: no evidence for a constancy of central conduction delays," *Neuroscience letters*,

[11] S. Mann, T. Furness, Y. Yuan, J. Iorio, and Z. Wang, "All reality: Virtual, augmented, mixed (x), mediated (x, y), and multimediated reality," *arXiv preprint arXiv:1804.08386*, 2018.

vol. 166, no. 1, p. 9—12, January 1994. [Online]. Available: https://doi.org/10.1016/0304-3940(94)90828-1

[4] J. Peddie, "Technology issues," in *Augmented Reality*. Springer, 2017, pp. 183–289.

[5] M. Clynes and N. Kline, "Cyborgs and space," *Astronautics*, vol. 14, no. 9, pp. 26–27,and 74–75, Sept. 1960.

[6] S. Mann, "Wearable computing: Toward humanistic intelligence," *Intelligent Systems, IEEE*, vol. 16, pp. 10–15, 2001.

[7] J. Benditt, "Humachines," vol. 102, Issue 3, pp. Cover + 36–41.

[8] M. Minsky, R. Kurzweil, and S. Mann, "Society of intelligent veillance," in *IEEE ISTAS 2013*, pp. 13–17.

[9] S. Mann, "Telematic tubs against terror: Bathing in the immersive interactive media of the post-cyborg age," *Leonardo*, vol. 37, no. 5, pp. 372–373, 2004.

[10] S. Mann, M. L. Hao, and J. Warner, "Virtual reality games in sensory deprivation tanks," in *2018 IEEE Games, Entertainment, Media Conference (GEM)*. IEEE, 2018, pp. 1–9.

[12] D. Benarab, T. Napoléon, A. Alfalou, A. Verney, and P. Hellard, "Optimized swimmer tracking system by a dynamic fusion of correlation and color histogram techniques," *Optics communications*, vol. 356, pp. 256–268, 2015.

[13] ——, "Optimized swimmer tracking system based on a novel multi-related-targets approach," *Optics and lasers in engineering*, vol. 89, pp. 195–202, 2017.